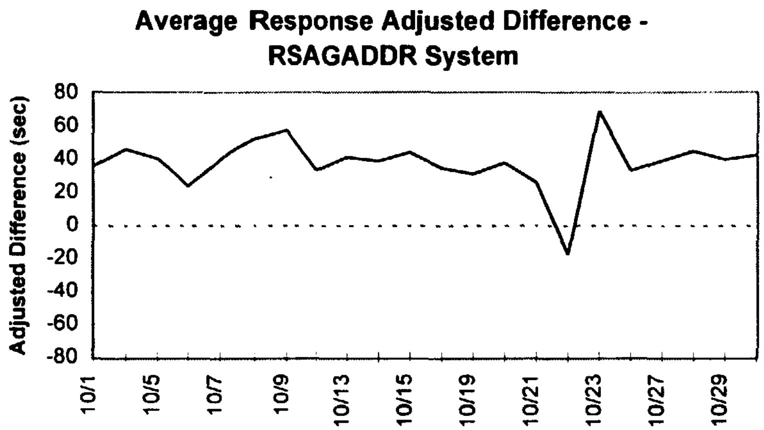
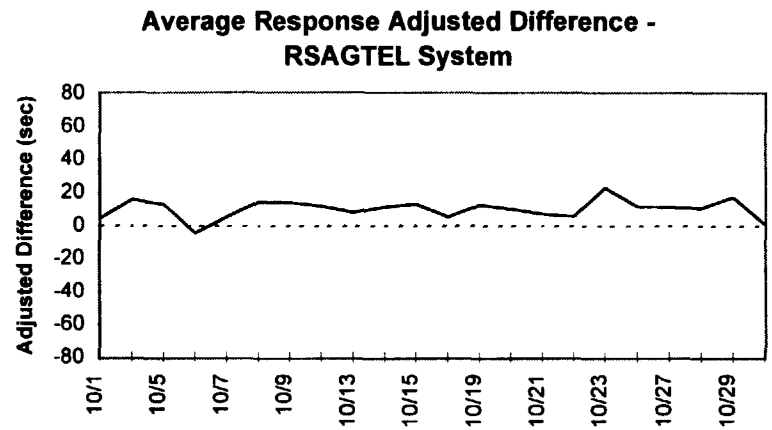
October 1998
OSS Response Interval -- Decision Page





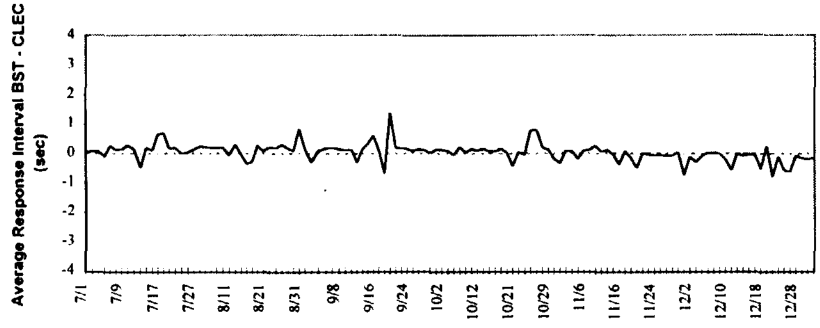
Mean	Standard Error	Test Statistic	df	P-value (percent)
38.26	6.64	5.76	21	0.00

Mean	Standard Error	Test Statistic	df	P-value (percent)
10.15	5.22	1.95	21	3.26

Note: On weekdays for which CLEC observations existed, but BST observations were handled as missing values and values were imputed.

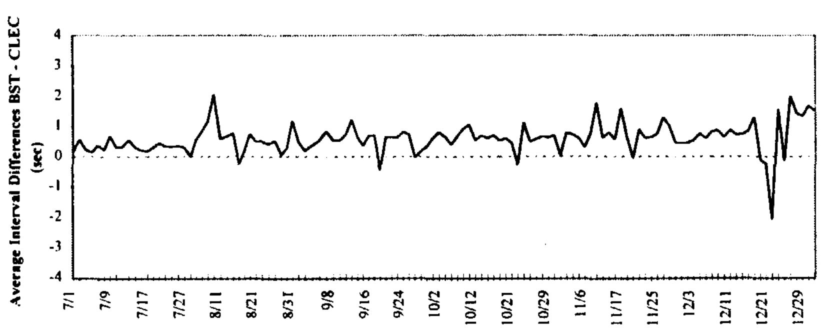
July - December 1998
OSS Response Interval -- Document Page

Average OSS Response Interval Differences - ATLAS System



Average OSS Interval Differences - DSAP System

Average OSS Interval Differences - RSAGADDR System



Average OSS Interval Differences - RSAGTEL System

July - December 1998
OSS Response Interval -- Document Page

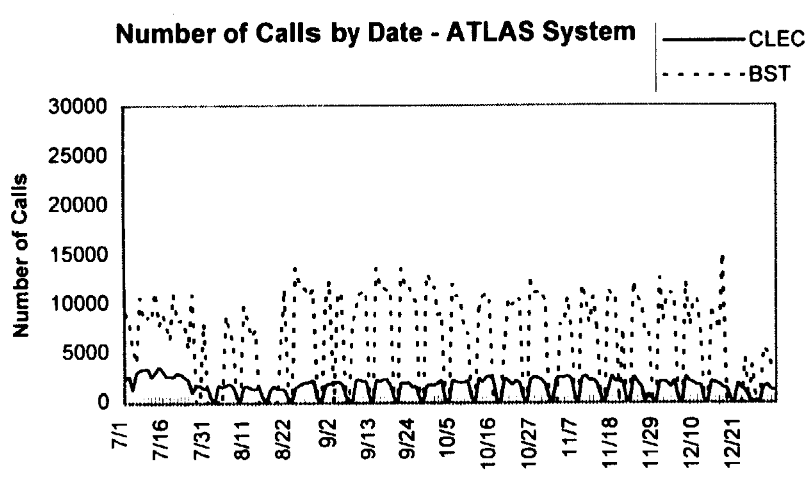
Time Series Analysis Results

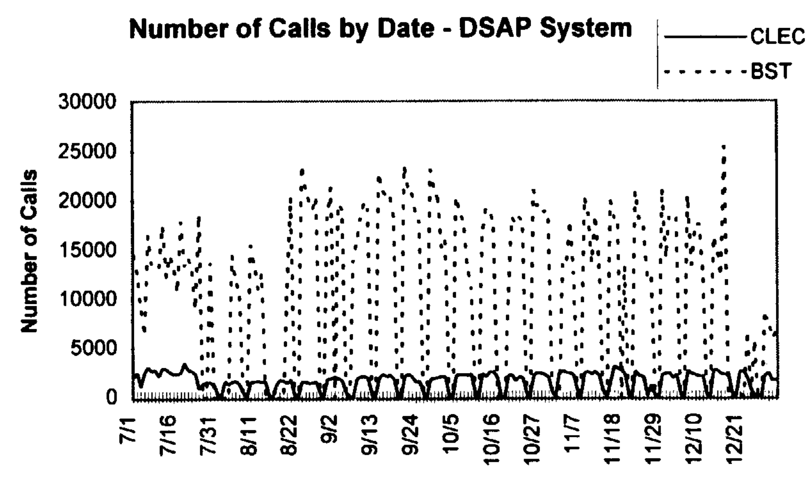
System	Result	Parameters	Estimated White Noise Variance
ATLAS	AR(1) model	ϕ_1=.168682	1504955000
DSAP	AR(3) model	ϕ_1=.075329 ϕ_2=-.086617 ϕ_3=-.256442	353610000
RSAGADDR	AR(3) model	ϕ_1=.173137 ϕ_2=-.049263 ϕ_3=.173076	518551000
RSAGTEL	white noise	-	599188000

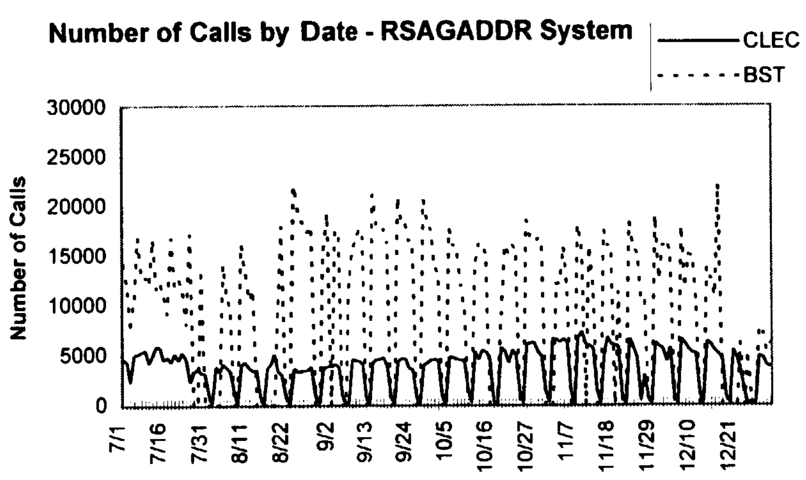
The residuals of each series were tested under the Ljung-Box and McLeod-Li portmanteau tests of independence. These tests of independence assume independent data under the null hypothesis and are approximately chi-squared with twenty degrees of freedom.

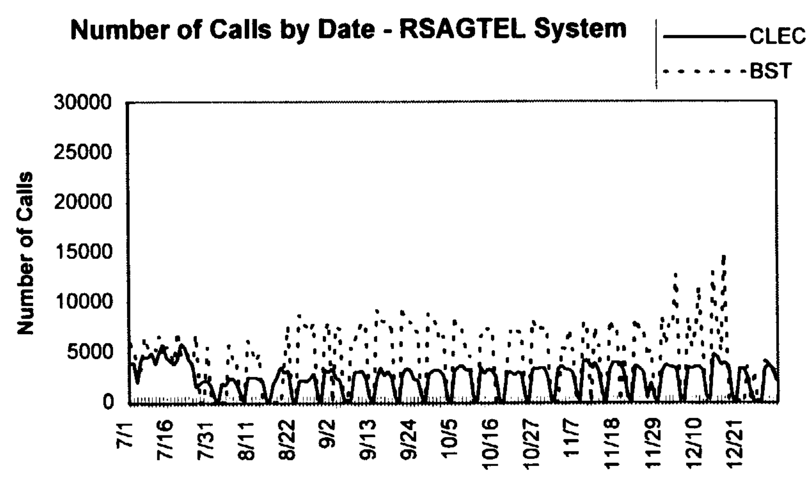
System	Ljung-Box test statistic	P-value (percent)	McLeod-Li test statistic	P-value (percent)
ATLAS	37.0630	1.1500	17.3750	62.8506
DSAP	23.1860	27.9754	18.2660	56.9890
RSAGADDR	15.5880	74.1833	15.5860	74.1953
RSAGTEL	12.8280	88.4641	2.6927	99.9998

July - December 1998
OSS Response Interval -- Document Page









October 1998
OSS Response Interval Imputed and Removed Values

Weekdays for which no BST value was reported, an adjusted difference was imputed for the purposes of time series analysis, thus handling the observation as if it were missing data. As a result of low volume of calls, all CLEC weekend values were removed.

CLEC Values Imputed (weekdays)			
Date	System	# calls	Average Seconds
10/16/98	ATLAS	2119	0.64
10/16/98	DSAP	1985	0.24
10/16/98	RSAGADDR	4693	1.11
10/16/98	RSAGTEL	2585	0.85

CLEC Values Removed (weekends)			
Date	System	# calls	Average Seconds
10/3/98	ATLAS	188	498.80
10/3/98	DSAPDDI	318	199.37
10/3/98	RSAGADDR	862	1154.44
10/3/98	RSAGTEL	520	677.08
10/4/98	ATLAS	37	812.59
10/4/98	DSAPDDI	109	168.39
10/4/98	RSAGADDR	146	1307.32
10/4/98	RSAGTEL	140	993.91
10/10/98	ATLAS	462	447.45
10/10/98	DSAPDDI	505	180.15
10/10/98	RSAGADDR	1408	873.42
10/10/98	RSAGTEL	576	644.02
10/11/98	ATLAS	14	998.86
10/11/98	DSAPDDI	108	661.96
10/11/98	RSAGADDR	181	1003.99
10/11/98	RSAGTEL	201	883.03
10/17/98	ATLAS	69	518.12
10/17/98	DSAPDDI	58	83824.45
10/17/98	RSAGADDR	262	1082.75
10/17/98	RSAGTEL	151	867.03

October 1998 - continued
OSS Response Interval Imputed and Removed Values

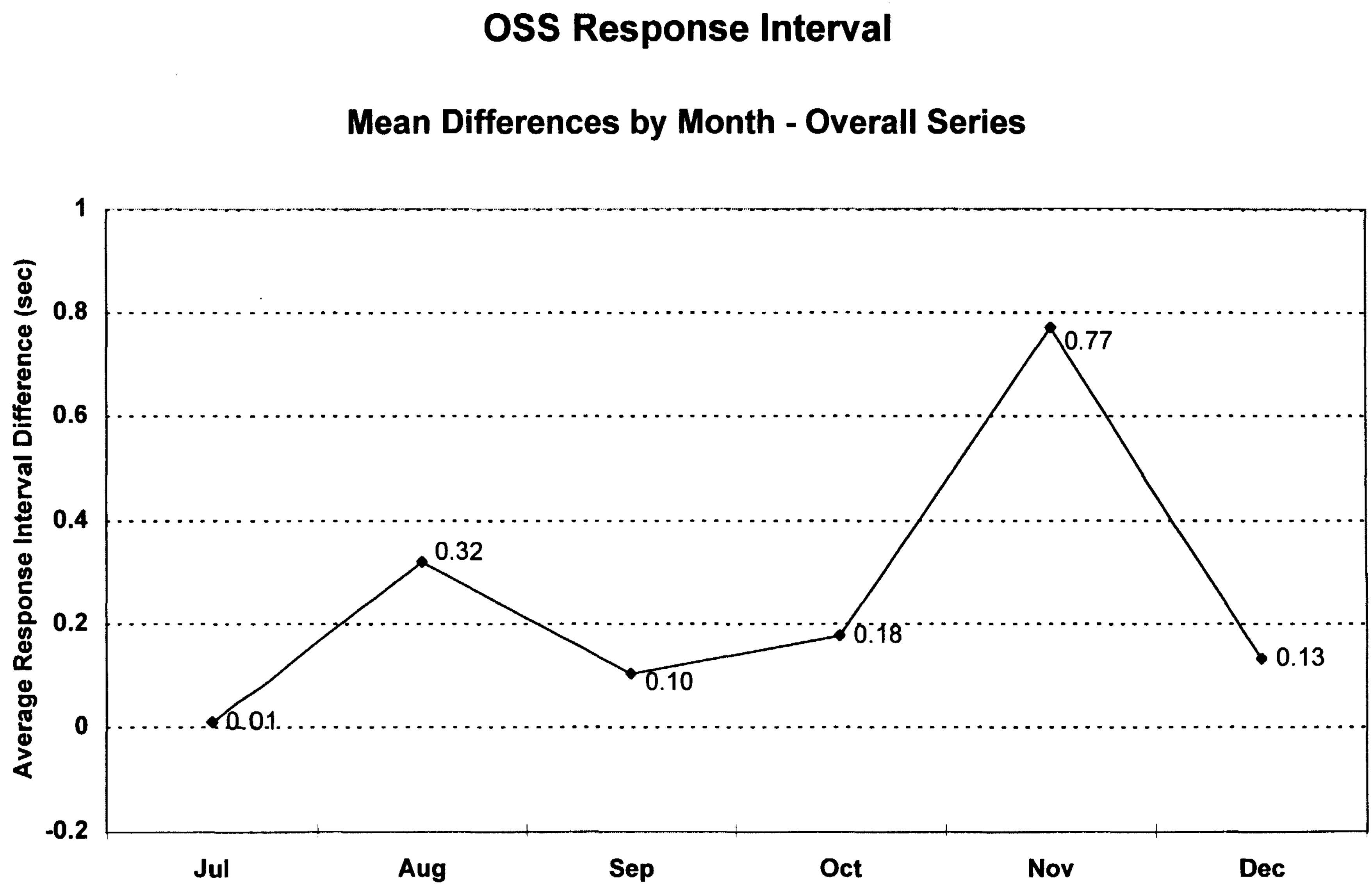
Weekdays for which no BST value was reported, an adjusted difference was imputed for the purposes of time series analysis, thus handling the observation as if it were missing data. As a result of low volume of calls, all CLEC weekend values were removed.

CLEC Values Imputed (weekdays)			
Date	**System**	**# calls**	**Average Seconds**

CLEC Values Removed (weekends)			
Date	**System**	**# calls**	**Average Seconds**
10/18/98	ATLAS	26	612.81
10/18/98	DSAPDDI	48	219.48
10/18/98	RSAGADDR	52	1501.17
10/18/98	RSAGTEL	33	1619.03
10/24/98	ATLAS	362	563.92
10/24/98	DSAPDDI	388	172.96
10/24/98	RSAGADDR	1280	883.35
10/24/98	RSAGTEL	510	642.36
10/25/98	ATLAS	68	585.31
10/25/98	DSAPDDI	95	144.23
10/25/98	RSAGADDR	146	1007.03
10/25/98	RSAGTEL	73	1250.78
10/31/98	ATLAS	487	492.71
10/31/98	DSAPDDI	504	160.71
10/31/98	RSAGADDR	1403	1548.30
10/31/98	RSAGTEL	616	947.04

OSS Response Interval

November 1998


OSS Response Interval
Mean Differences by Month - Overall Series
Average Response Interval Difference (sec)
1
0.8
0.6
0.4
0.2
0
-0.2
0.01
0.32
0.10
0.18
0.77
0.13
Jul
Aug
Sep
Oct
Nov
Dec

November 1998
OSS Response Interval Web Report

This table displays BellSouth and CLEC information based on data received. No official web report was available at time of report generation.

	BST (RNS)				CLEC (LENS)			
SYSTEM	< 2.3 SECONDS	> 6 SECONDS	AVERAGE SECONDS	# OF CALLS	< 2.3 SECONDS	> 6 SECONDS	AVERAGE SECONDS	# OF CALLS
RSAG								
BY TN			1.18	119848			0.88	72141
BY ADDR			1.98	261455			1.16	124034
ATLAS			0.69	181112			0.67	43265
DSAP			0.59	153195			0.36	48934
CRSACCTS[1]								
OASISNET[3]								
OASISBSN[2]								
OASISCAR[2]								
OASISLPC[2]								
OASISMTN[2]								
OASISOCP[3]								
OASISBIG[2]								
HAL/CRIS[2]								
COFI/USOC[2]								
PSIMS/ORB[2]								

Note 1: CSR data is retrieved via the CRSACCTS contract in RNS and the HAL/CRIS contract in LENS. The HAL/CRIS response time shown above includes processing time for filtering and formatting CSR data which is not included in the CRSACCTS contract. RNS time reflects the handling of residence orders only, while LENS time reflects the handling of both residence and more complex business orders.

Note 2: Service/feature availability is retrieved via a series of OASIS contracts in RNS and via calls to COFFI and P/SIMS in LENS.

Note 3: OASIS contract in RNS replaced by OASISBIG.

November 1998
OSS Response Interval Web Report Verification

This section reports whether there are any inconsistencies between the data reported on the web and calculations done during our processing. Excludes all non-like-to-like systems. Since individual calls are not reported, only Average Seconds and # of Calls can be verified.

Verify CLEC Web Report Values

Matched	Did Not Match

Verify BST Web Report Values

Matched	Did Not Match

CLEC Inconsistencies

	Average Seconds		# of Calls	
System	Web Report	Raw Data	Web Report	Raw Data

BST Inconsistencies

	Average Seconds		# of Calls	
System	Web Report	Raw Data	Web Report	Raw Data

Explanation (as needed): Official web report was not available.

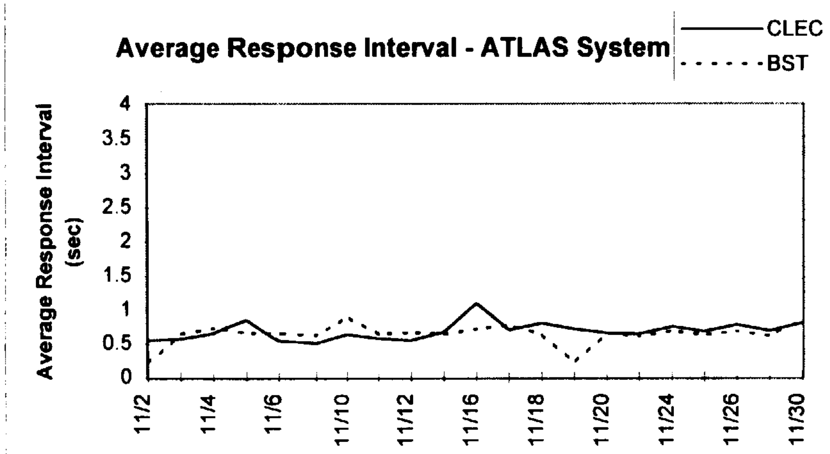
November 1998
OSS Response Interval Filtering Information

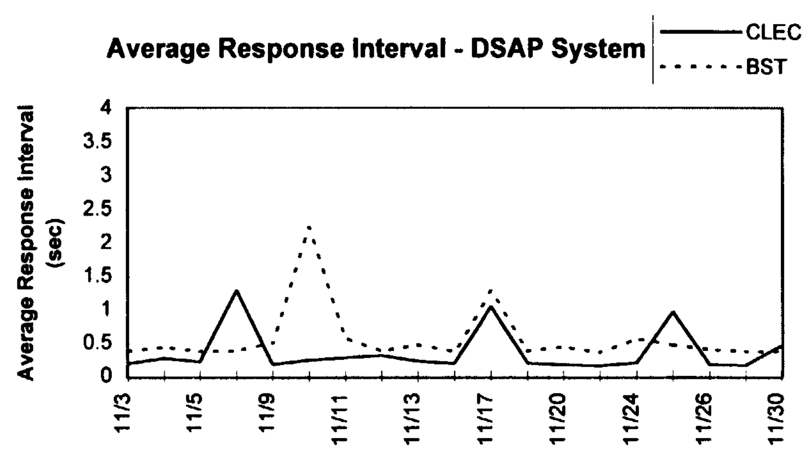
This table displays information about the size of the database files and the cases that were removed from the analysis.

BST	
Unfiltered Total	**193**
Uncomparable System Records Removed	
CRSACCTS	**21**
OASISBSN	**21**
OASISCAR	**18**
OASISLPC	**18**
OASISMTN	**18**
OASISBIG	**19**
Weekend Records Removed	**0**
FILTERED TOTAL	**78**

CLEC	
Unfiltered Total	**209**
Uncomparable System Records Removed	
HAL/CRIS	**30**
COFI/USOC	**29**
PSIMS/ORB	**30**
Weekend Records Removed	**36**
FILTERED TOTAL	**84**

November 1998
OSS Response Interval -- Descriptive Page



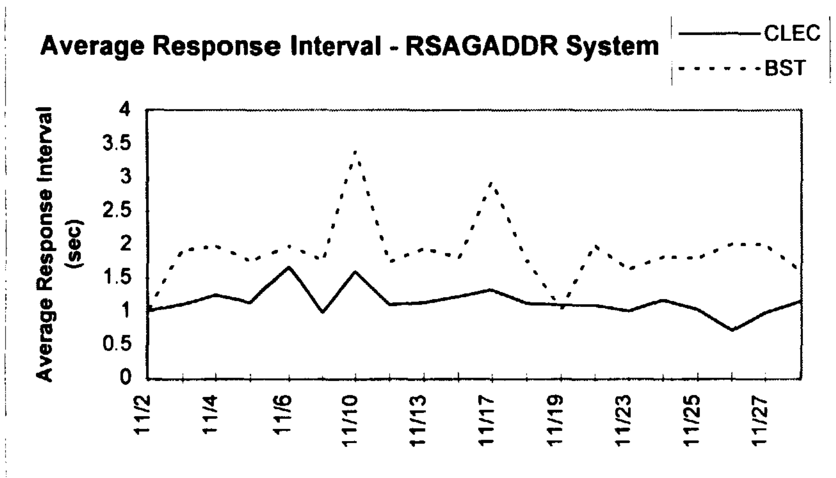


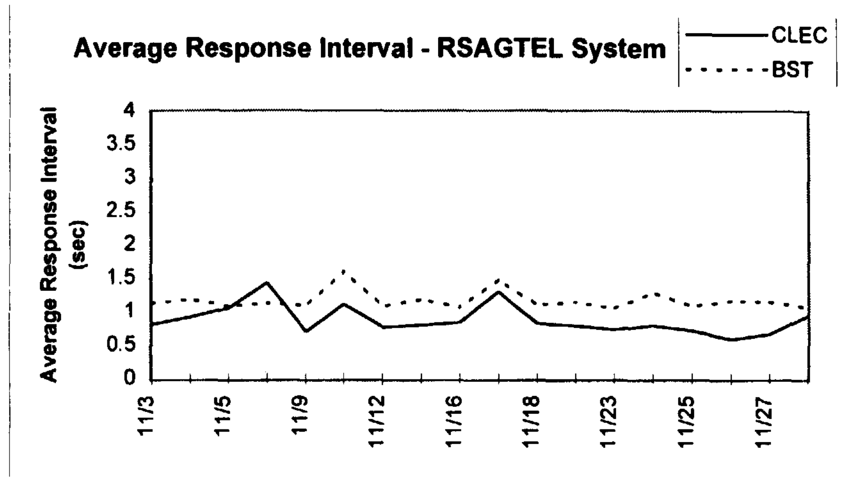
Service Provider	Mean
BST	0.6877
CLEC	0.6709
Difference	0.0168

Service Provider	Mean
BST	0.5852
CLEC	0.3680
Difference	0.2172

Note: Sums of squares were unavailable for the Response Interval, thus a standard deviation could not be calculated for these measures.

November 1998
OSS Response Interval -- Descriptive Page



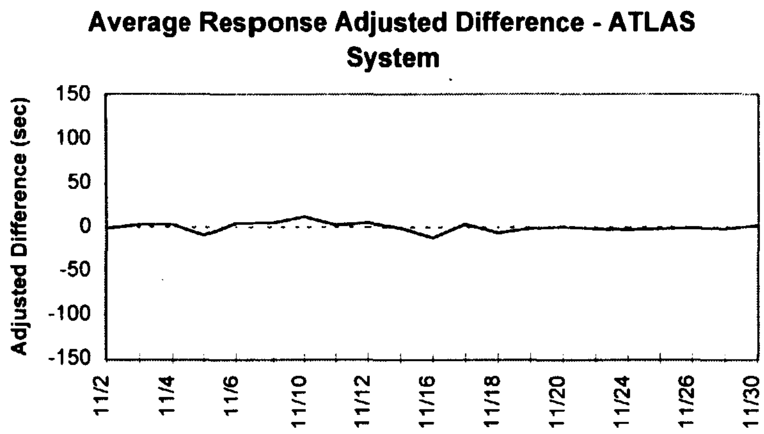


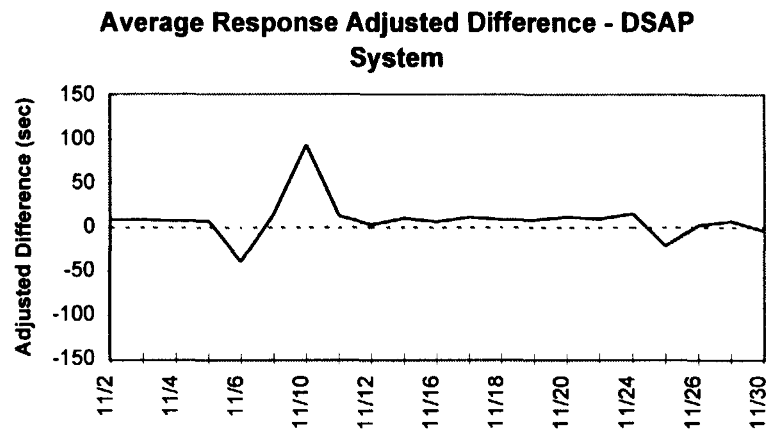
Service Provider	Mean
BST	1.9755
CLEC	1.1698
Difference	0.8057

Service Provider	Mean
BST	1.1762
CLEC	0.8868
Difference	0.2894

Note: Sums of squares were unavailable for the Response Interval, thus a standard deviation could not be calculated for these measures.

November 1998
OSS Response Interval -- Decision Page



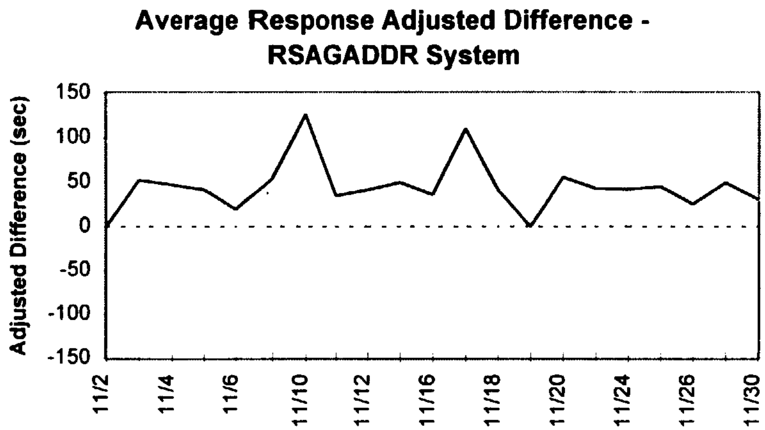


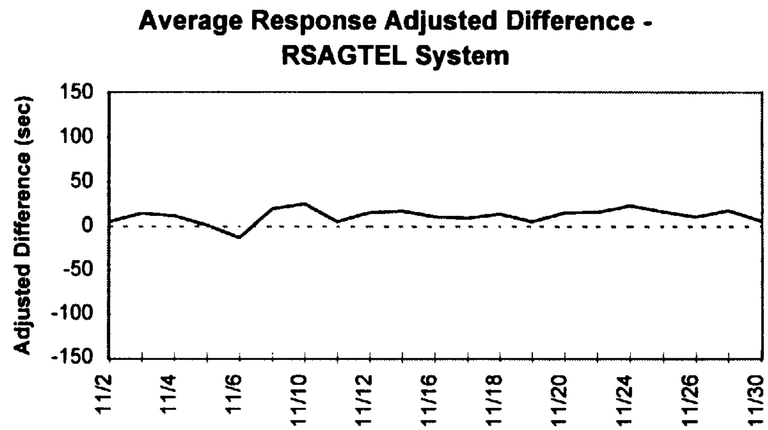
Mean	Standard Error	Test Statistic	df	P-value (percent)
-0.09	2.67	-0.03	20	48.72

Mean	Standard Error	Test Statistic	df	P-value (percent)
9.06	3.34	2.71	20	0.67

Note: On weekdays for which CLEC observations existed, but BST observations were handled as missing values and values were imputed.

November 1998
OSS Response Interval -- Decision Page





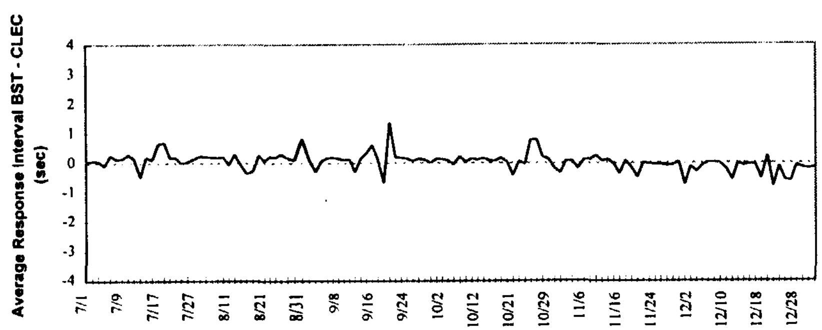
Mean	Standard Error	Test Statistic	df	P-value (percent)
4.36	6.79	0.64	20	26.38

Mean	Standard Error	Test Statistic	df	P-value (percent)
11.86	5.34	2.22	20	1.91

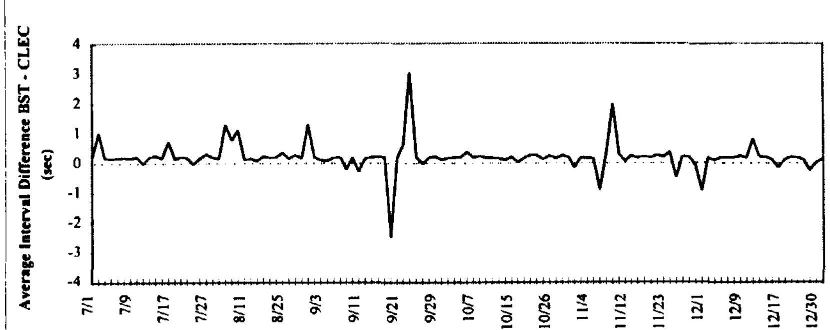
Note: On weekdays for which CLEC observations existed, but BST observations did not, the BST observations were handled as missing values and values were imputed.

July - December 1998
OSS Response Interval -- Document Page

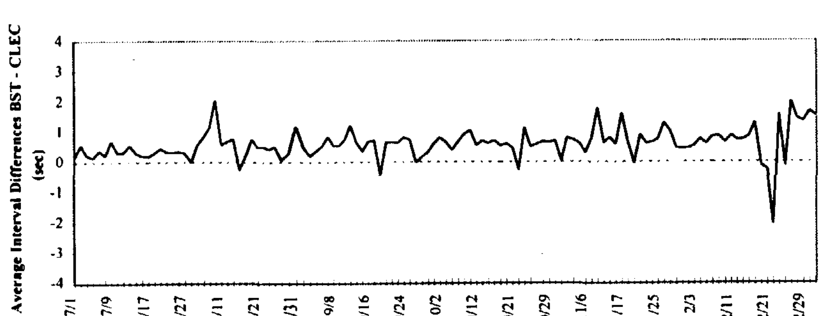
Average OSS Response Interval Differences - ATLAS System



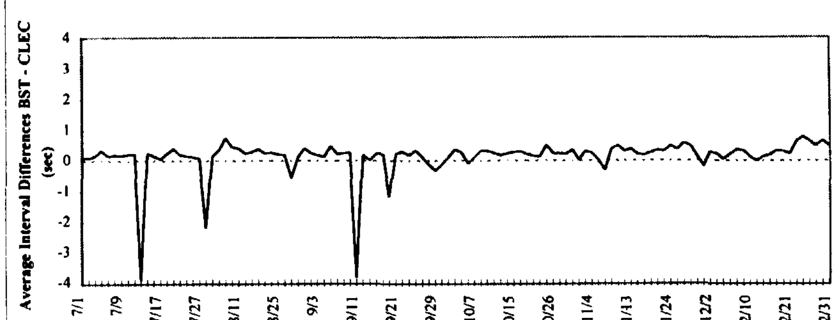
Average OSS Interval Differences - DSAP System



Average OSS Interval Differences - RSAGADDR System



Average OSS Interval Differences - RSAGTEL System



July - December 1998
OSS Response Interval -- Document Page

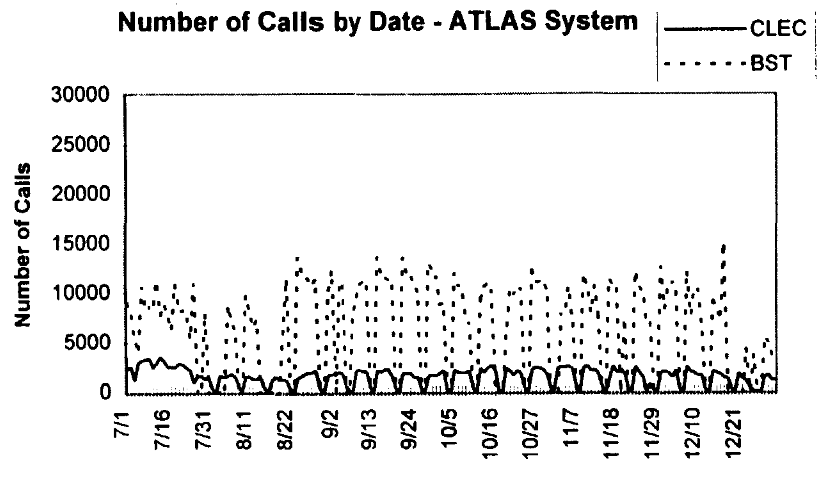
Time Series Analysis Results

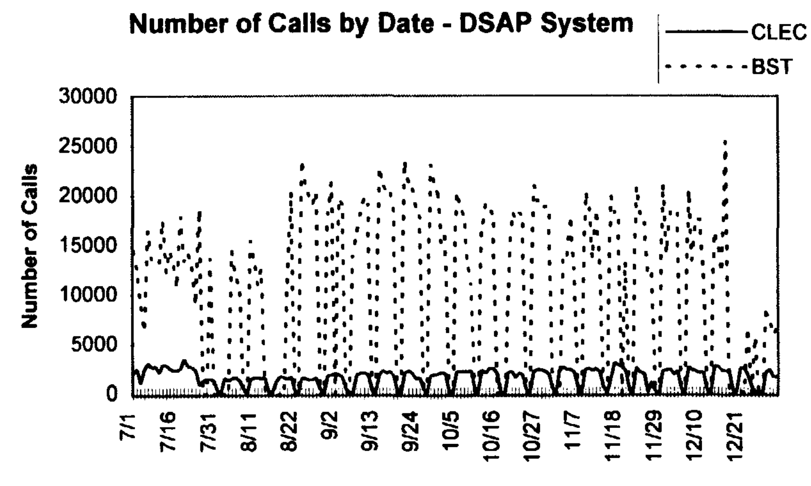
System	Result	Parameters	Estimated White Noise Variance
ATLAS	AR(1) model	ϕ_1=.168682	1504955000
DSAP	AR(3) model	ϕ_1=.075329 ϕ_2=-.086617 ϕ_3=-.256442	353610000
RSAGADDR	AR(3) model	ϕ_1=.173137 ϕ_2=-.049263 ϕ_3=.173076	518551000
RSAGTEL	white noise	-	599188000

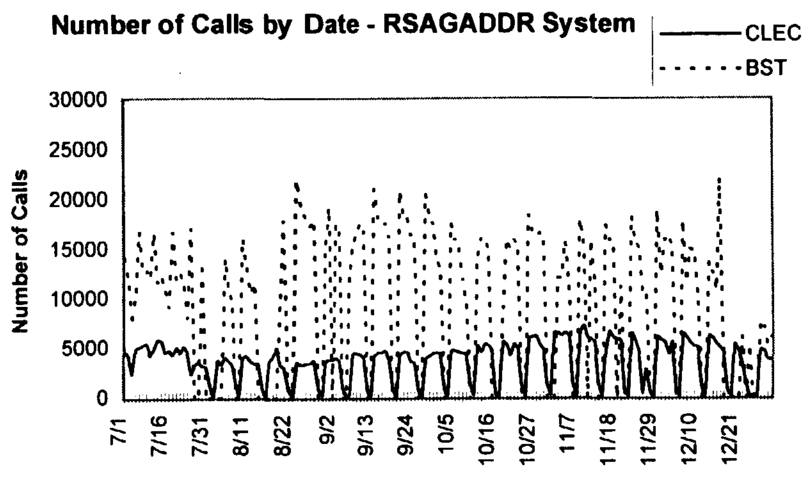
The residuals of each series were tested under the Ljung-Box and McLeod-Li portmanteau tests of independence. These tests of independence assume independent data under the null hypothesis and are approximately chi-squared with twenty degrees of freedom.

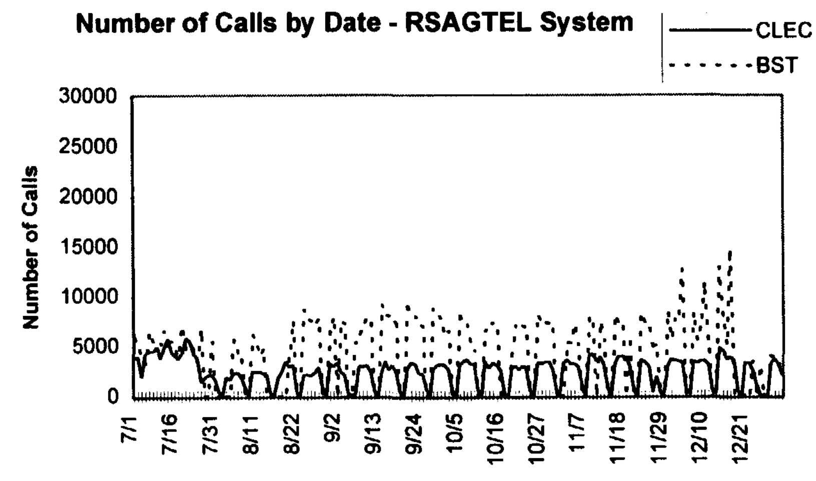
System	Ljung-Box test statistic	P-value (percent)	McLeod-Li test statistic	P-value (percent)
ATLAS	37.0630	1.1500	17.3750	62.8506
DSAP	23.1860	27.9754	18.2660	56.9890
RSAGADDR	15.5880	74.1833	15.5860	74.1953
RSAGTEL	12.8280	88.4641	2.6927	99.9998

July - December 1998
OSS Response Interval -- Document Page









November 1998
OSS Response Interval Imputed and Removed Values

Weekdays for which no BST value was reported, an adjusted difference was imputed for the purposes of time series analysis, thus handling the observation as if it were missing data. As a result of low volume of calls, all CLEC weekend values were removed.

CLEC Values Imputed (weekdays)

Date	System	# calls	Average Seconds
11/2/98	DSAP	2395	0.18
11/2/98	RSAGTEL	3302	0.74
11/11/98	RSAGADDR	6011	1.05
11/11/98	RSAGTEL	3480	0.81
11/19/98	DSAP	2713	0.18
11/19/98	RSAGTEL	3809	0.79

CLEC Values Removed (weekends)

Date	System	# calls	Average Seconds
11/1/98	ATLAS	135	526.79
11/1/98	DSAPDDI	142	151.55
11/1/98	RSAGADDR	239	1242.12
11/1/98	RSAGTEL	82	911.23
11/7/98	ATLAS	442	477.39
11/7/98	DSAPDDI	606	169.22
11/7/98	RSAGADDR	1406	748.73
11/7/98	RSAGTEL	1117	478.39
11/8/98	ATLAS	8	760.50
11/8/98	DSAPDDI	18	248.61
11/8/98	RSAGADDR	47	1830.79
11/8/98	RSAGTEL	36	1343.61
11/14/98	ATLAS	563	595.42
11/14/98	DSAPDDI	679	218.41
11/14/98	RSAGADDR	1481	1003.00
11/14/98	RSAGTEL	811	707.99
11/15/98	ATLAS	33	954.61
11/15/98	DSAPDDI	67	176.94
11/15/98	RSAGADDR	93	1548.32
11/15/98	RSAGTEL	81	1207.64

November 1998 - continued
OSS Response Interval Imputed and Removed Values

Weekdays for which no BST value was reported, an adjusted difference was imputed for the purposes of time series analysis, thus handling the observation as if it were missing data. As a result of low volume of calls, all CLEC weekend values were removed.

CLEC Values Imputed (weekdays)

Date	System	# calls	Average Seconds

CLEC Values Removed (weekends)

Date	System	# calls	Average Seconds
11/21/98	ATLAS	379	644.61
11/21/98	DSAPDDI	514	181.26
11/21/98	RSAGADDR	978	899.84
11/21/98	RSAGTEL	712	621.08
11/22/98	ATLAS	2	1163.00
11/22/98	DSAPDDI	8	196.13
11/22/98	RSAGADDR	42	1541.64
11/22/98	RSAGTEL	20	847.05
11/28/98	ATLAS	182	717.25
11/28/98	DSAPDDI	170	177.15
11/28/98	RSAGADDR	577	983.71
11/28/98	RSAGTEL	254	755.45
11/29/98	ATLAS	18	1146.78
11/29/98	DSAPDDI	30	216.40
11/29/98	RSAGADDR	55	1224.29
11/29/98	RSAGTEL	284	686.68

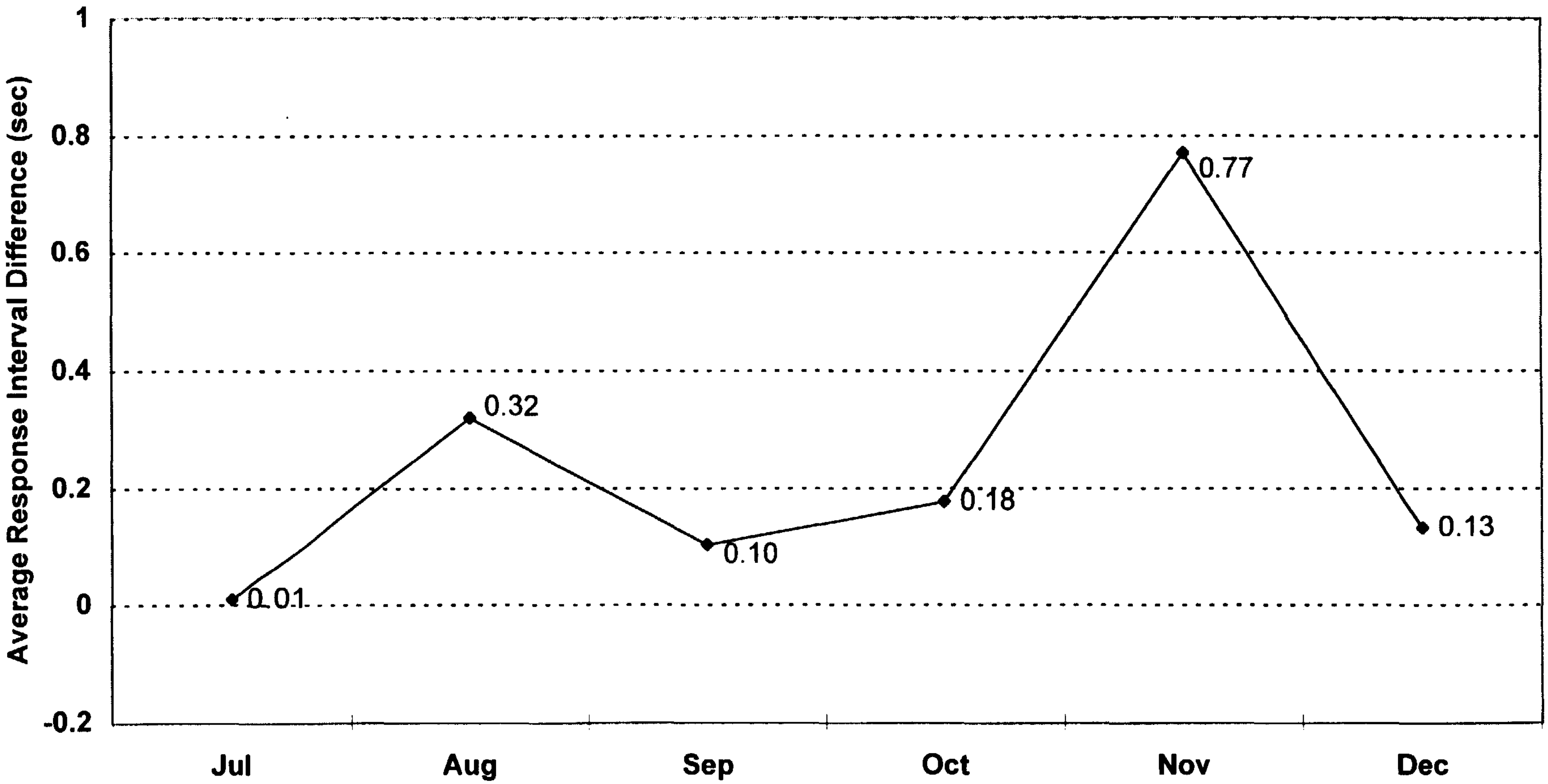
BELLSOUTH

OSS Response Interval

December 1998

OSS Response Interval

Mean Differences by Month - Overall Series



December 1998
OSS Response Interval Web Report

This table displays BellSouth and CLEC information as downloaded from the December monthly web reports, https://clec.bellsouth.com

	BST (RNS)				CLEC (LENS)			
SYSTEM	< 2.3 SECONDS	> 6 SECONDS	AVERAGE SECONDS	# OF CALLS	< 2.3 SECONDS	> 6 SECONDS	AVERAGE SECONDS	# OF CALLS
RSAG								
BY TN	97.04%	1.21%	1.10	132351	97.92%	0.41%	0.88	76739
BY ADDR	92.12%	2.20%	2.09	226167	92.60%	1.08%	1.27	113261
ATLAS	97.62%	1.24%	0.67	153195	97.87%	0.27%	0.83	36318
DSAP	98.69%	0.67%	0.44	259273	99.47%	0.26%	0.39	49782
CRSACCTS[1]	95.27%	1.42%	1.55	665664				
OASISNET[3]								
OASISBSN[2]	98.26%	0.56%	0.66	855375				
OASISCAR[2]	96.81%	0.49%	0.87	408897				
OASISLPC[2]	97.85%	0.51%	0.86	150745				
OASISMTN[2]	98.73%	0.48%	0.74	264830				
OASISOCP[3]								
OASISBIG[2]	3.34%	24.43%	5.28	713391				
HAL/CRIS[2]					5.14%	30.51%	6.08	646481
COFI/USOC[2]					99.51%	0.14%	0.42	15510
PSIMS/ORB[2]					56.47%	11.28%	2.53	42039

Note 1: CSR data is retrieved via the CRSACCTS contract in RNS and the HAL/CRIS contract in LENS. The HAL/CRIS response time shown above includes processing time for filtering and formatting CSR data which is not included in the CRSACCTS contract. RNS time reflects the handling of residence orders only, while LENS time reflects the handling of both residence and more complex business orders.

Note 2: Service/feature availability is retrieved via a series of OASIS contracts in RNS and via calls to COFFI and P/SIMS in LENS.

Note 3: OASIS contract in RNS replaced by OASISBIG.

December 1998
OSS Response Interval Web Report Verification

This section reports whether there are any inconsistencies between the data reported on the web and calculations done during our processing. Excludes all non-like-to-like systems. Since individual calls are not reported, only Average Seconds and # of Calls can be verified.

Verify CLEC Web Report Values

Matched	Did Not Match
8	0

Verify BST Web Report Values

Matched	Did Not Match
8	0

CLEC Inconsistencies

	Average Seconds		# of Calls	
System	Web Report	Raw Data	Web Report	Raw Data

BST Inconsistencies

	Average Seconds		# of Calls	
System	Web Report	Raw Data	Web Report	Raw Data

Explanation (as needed)

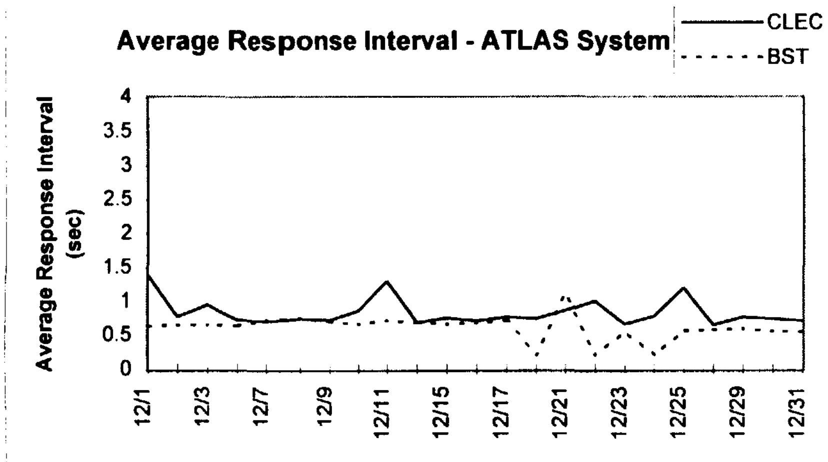
December 1998
OSS Response Interval Filtering Information

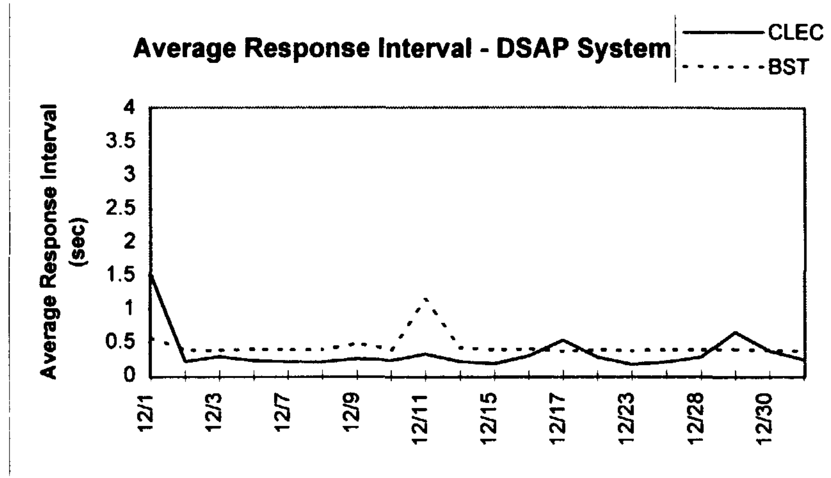
This table displays information about the size of the database files and the cases that were removed from the analysis.

BST	
Unfiltered Total	212
Uncomparable System Records Removed	
CRSACCTS	**23**
OASISBSN	**23**
OASISCAR	**20**
OASISLPC	**20**
OASISMTN	**20**
OASISBIG	**20**
Weekend Records Removed	**0**
FILTERED TOTAL	**86**

CLEC	
Unfiltered Total	214
Uncomparable System Records Removed	
HAL/CRIS	**31**
COFI/USOC	**29**
PSIMS/ORB	**30**
Weekend Records Removed	**32**
FILTERED TOTAL	**92**

December 1998
OSS Response Interval -- Descriptive Page



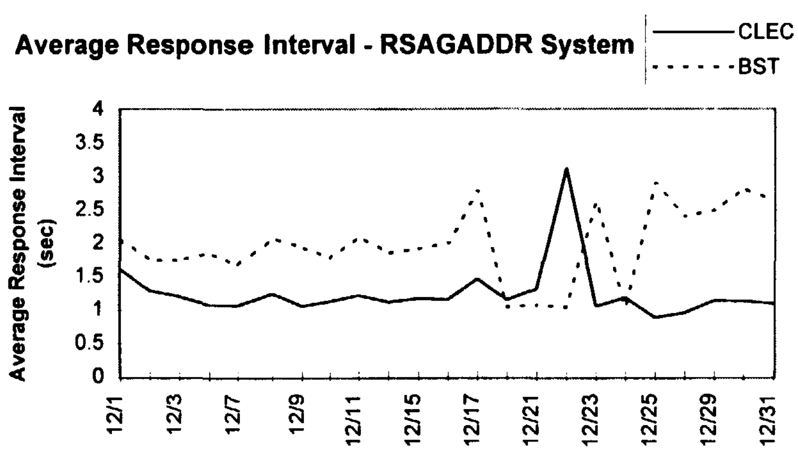


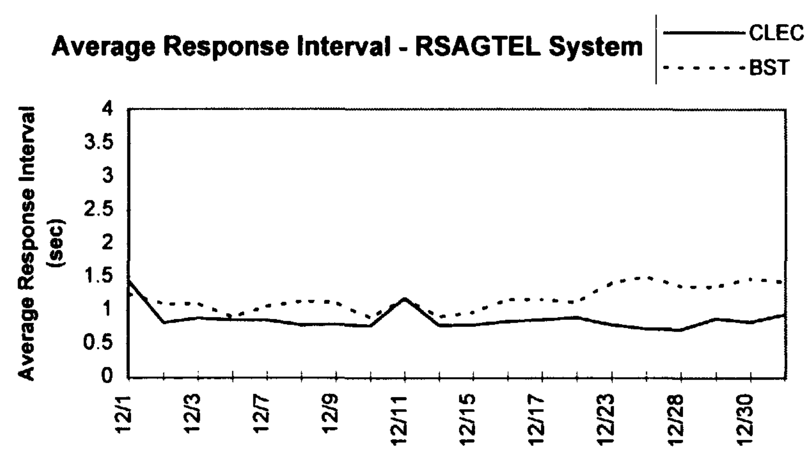
Service Provider	Mean
BST	0.6695
CLEC	0.8303
Difference	-0.1608

Service Provider	Mean
BST	0.4403
CLEC	0.3987
Difference	0.0416

Note: Sums of squares were unavailable for the Response Interval, thus a standard deviation could not be calculated for these measures.

December 1998
OSS Response Interval -- Descriptive Page



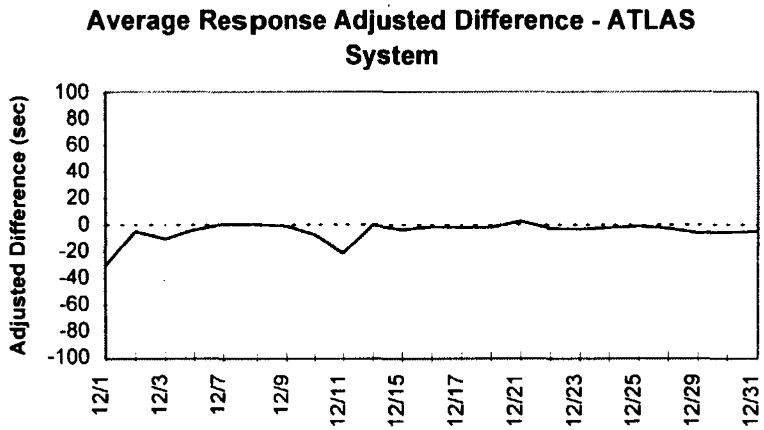


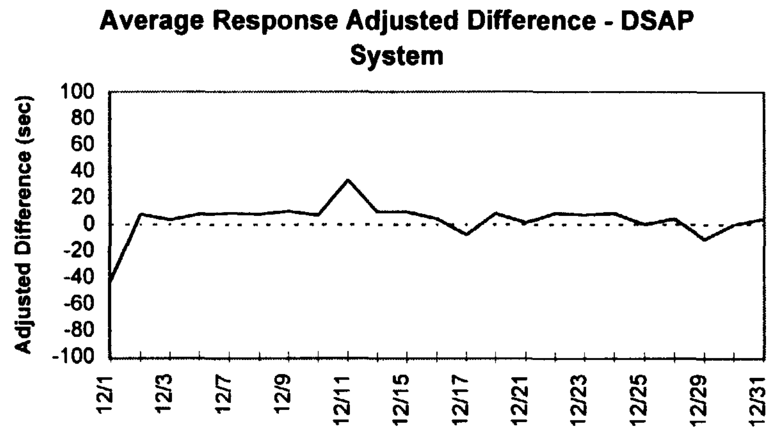
Service Provider	Mean
BST	2.0860
CLEC	1.2742
Difference	0.8118

Service Provider	Mean
BST	1.0995
CLEC	0.8857
Difference	0.2138

Note: Sums of squares were unavailable for the Response Interval, thus a standard deviation could not be calculated for these measures.

December 1998
OSS Response Interval -- Decision Page



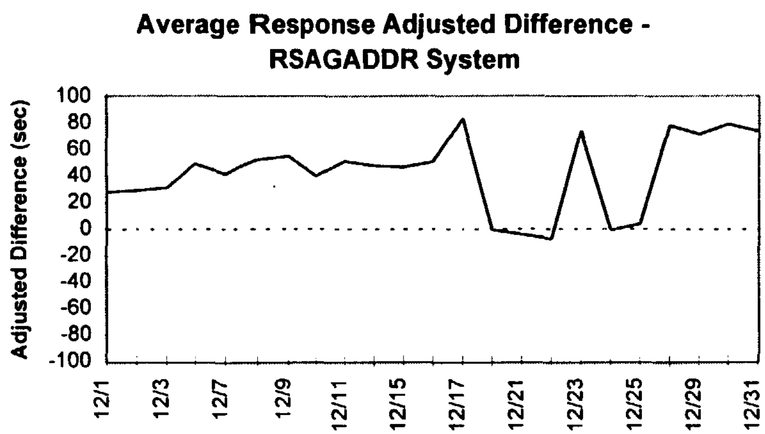


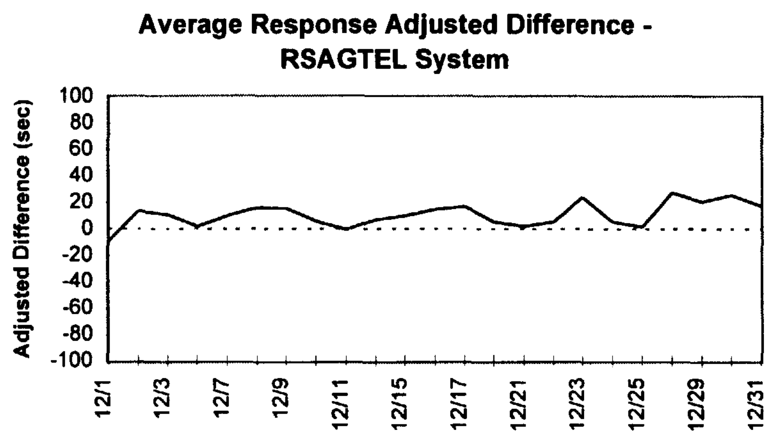
Mean	Standard Error	Test Statistic	df	P-value (percent)
-5.06	2.55	-1.98	22	3.01

Mean	Standard Error	Test Statistic	df	P-value (percent)
4.78	3.18	1.50	22	7.38

Note: On weekdays for which CLEC observations existed, but BST observations did not, the BST observations were handled as missing values and values were imputed.

December 1998
OSS Response Interval -- Decision Page





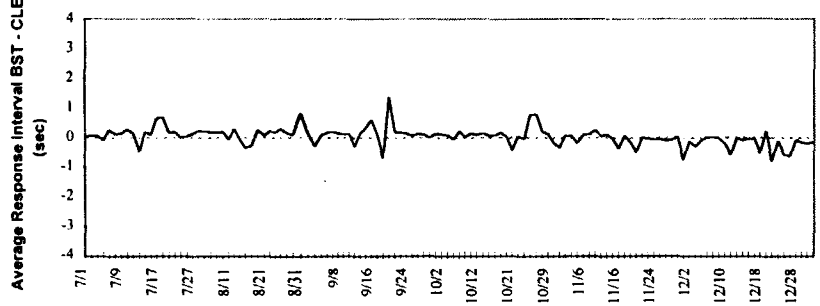
Mean	Standard Error	Test Statistic	df	P-value (percent)
43.34	6.50	6.66	22	0.00

Mean	Standard Error	Test Statistic	df	P-value (percent)
10.90	5.10	2.14	22	2.21

Note: On weekdays for which CLEC observations existed, but BST observations did not, the BST observations were handled as missing values and values were imputed.

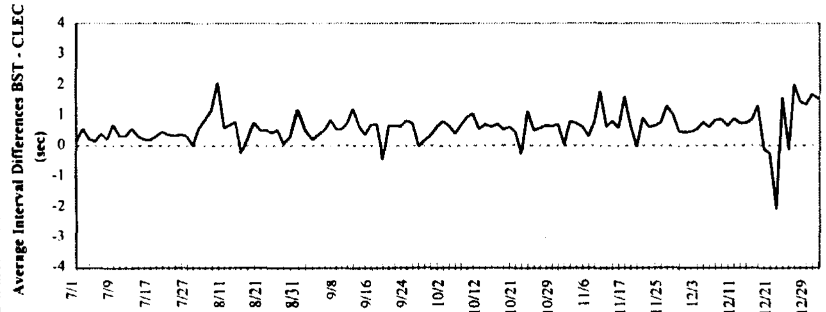
July - December 1998
OSS Response Interval -- Document Page

Average OSS Response Interval Differences - ATLAS System

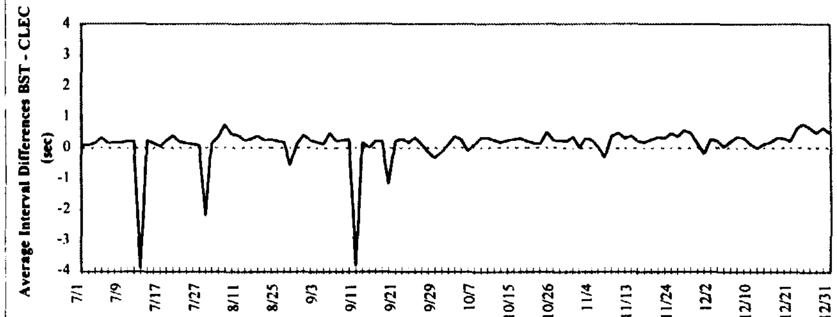


Average OSS Interval Differences - DSAP System

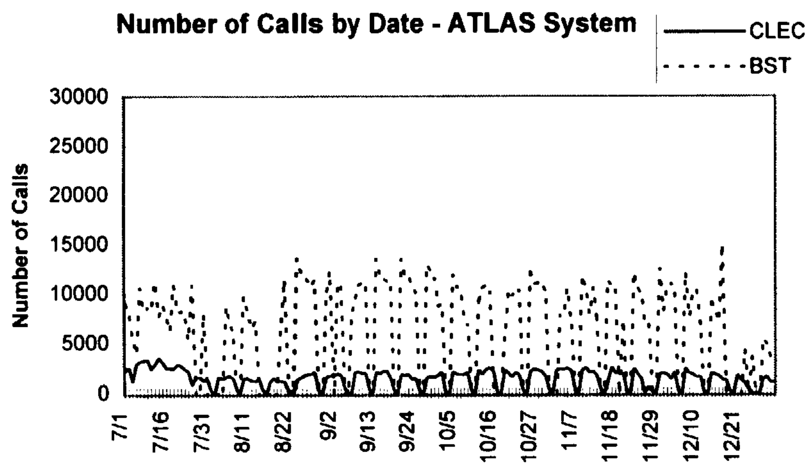
Average OSS Interval Differences - RSAGADDR System

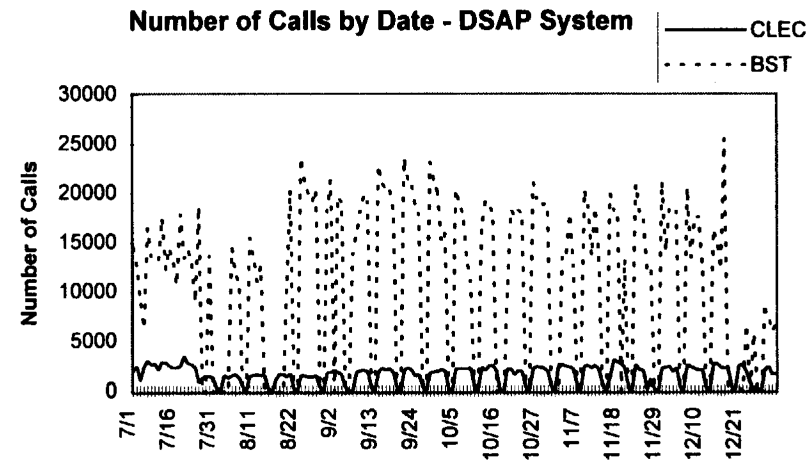


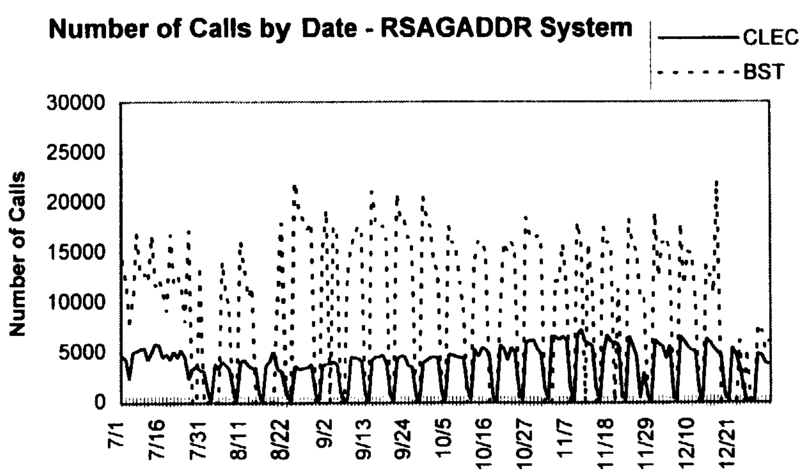
Average OSS Interval Differences - RSAGTEL System

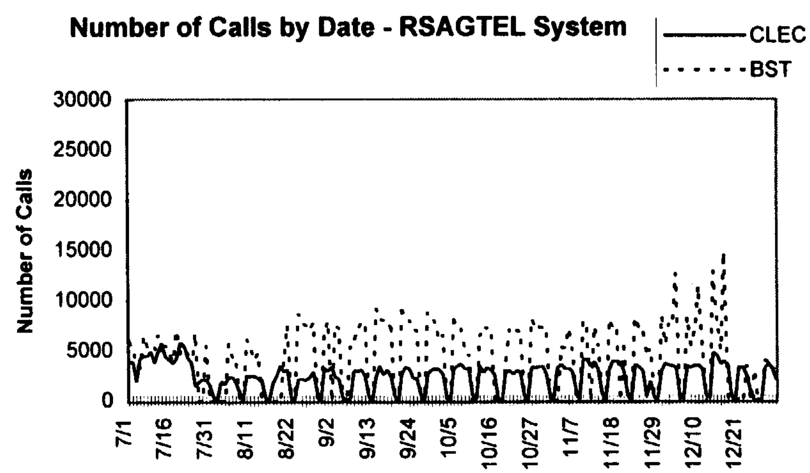


July - December 1998
OSS Response Interval -- Document Page









July - December 1998
OSS Response Interval -- Document Page

Time Series Analysis Results

System	Result	Parameters	Estimated White Noise Variance
ATLAS	AR(1) model	ϕ_1=.168682	1504955000
DSAP	AR(3) model	ϕ_1=.075329 ϕ_2=-.086617 ϕ_3=-.256442	353610000
RSAGADDR	AR(3) model	ϕ_1=.173137 ϕ_2=-.049263 ϕ_3=.173076	518551000
RSAGTEL	white noise	-	599188000

The residuals of each series were tested under the Ljung-Box and McLeod-Li portmanteau tests of independence. These tests of independence assume independent data under the null hypothesis and are approximately chi-squared with twenty degrees of freedom.

System	Ljung-Box test statistic	P-value (percent)	McLeod-Li test statistic	P-value (percent)
ATLAS	37.0630	1.1500	17.3750	62.8506
DSAP	23.1860	27.9754	18.2660	56.9890
RSAGADDR	15.5880	74.1833	15.5860	74.1953
RSAGTEL	12.8280	88.4641	2.6927	99.9998

December 1998
OSS Response Interval Imputed and Removed Values

Weekdays for which no BST value was reported, an adjusted difference was imputed for the purposes of time series analysis, thus handling the observation as if it were missing data. As a result of low volume of calls, all CLEC weekend values were removed.

CLEC Values Imputed (weekdays)

Date	System	# calls	Average Seconds
12/18/98	DSAP	2262	0.46
12/18/98	RSAGTEL	3567	0.79
12/22/98	DSAP	2615	1.13
12/22/98	RSAGTEL	3348	1.31
12/24/98	DSAP	821	0.18
12/24/98	RSAGTEL	990	0.77

CLEC Values Removed (weekends)

Date	System	# calls	Average Seconds
12/5/98	ATLAS	291	684.58
12/5/98	DSAPDDI	348	193.14
12/5/98	RSAGADDR	796	920.48
12/5/98	RSAGTEL	609	786.79
12/6/98	ATLAS	26	1063.08
12/6/98	DSAPDDI	172	160.51
12/6/98	RSAGADDR	254	1326.80
12/6/98	RSAGTEL	232	810.28
12/12/98	ATLAS	231	557.66
12/12/98	DSAPDDI	333	174.42
12/12/98	RSAGADDR	1019	792.91
12/12/98	RSAGTEL	585	646.73
12/13/98	ATLAS	11	1375.00
12/13/98	DSAPDDI	111	176.98
12/13/98	RSAGADDR	200	1406.60
12/13/98	RSAGTEL	262	871.61
12/19/98	ATLAS	263	591.35
12/19/98	DSAPDDI	323	167.71
12/19/98	RSAGADDR	723	974.50
12/19/98	RSAGTEL	441	657.06

December 1998 - continued
OSS Response Interval Imputed and Removed Values

Weekdays for which no BST value was reported, an adjusted difference was imputed for the purposes of time series analysis, thus handling the observation as if it were missing data. As a result of low volume of calls, all CLEC weekend values were removed.

CLEC Values Imputed (weekdays)			
Date	System	# calls	Average Seconds

CLEC Values Removed (weekends)			
Date	System	# calls	Average Seconds
12/20/98	ATLAS	34	924.76
12/20/98	DSAPDDI	121	171.21
12/20/98	RSAGADDR	152	1494.93
12/20/98	RSAGTEL	151	941.09
12/26/98	ATLAS	72	696.19
12/26/98	DSAPDDI	153	145.78
12/26/98	RSAGADDR	298	945.02
12/26/98	RSAGTEL	144	606.42
12/27/98	ATLAS	27	1107.07
12/27/98	DSAPDDI	49	178.31
12/27/98	RSAGADDR	138	1357.40
12/27/98	RSAGTEL	88	946.67

BellSouth Action Item 16-11,12,14
Follow-on Statistical Analysis
of
BellSouth Telecommunications, Inc.
Performance Measure Data

The data analysis and data presentation in this report include significant additions and improvements to the Interim Statistical Analysis Report, submitted to the Louisiana Public Service Commission, Docket U-22252, Subdocket C, on November 19, 1998. The changes in the presentation are to provide better documentation and to make the process as nearly self-documenting as possible. In the revised methodology, Ernst & Young has responded to concerns raised at the November workshop, and we have also incorporated additional improvements. The changes in the data analysis are outlined below; a more detailed description of each is then provided. The formula for each calculation is given last section.

Summary of Changes or Additions in the Data Analysis

1. Data Trimming – The FCC has suggested that a "general rule" for trimming the extreme tail of the observations is needed. We have provided one that trims the BST data more severely than in the previous analysis. This rule is used on the Order Completion Interval data.

2. Weighting to the BST Distribution – As requested, we now show the test computed by adjusting or weighting the CLEC observations to the BST distribution, as well as the original analysis which adjusts the BST data to the CLEC distribution.

3. Increasing Sensitivity of the BST Test to Inequality in Standard Deviations – We have made an adjustment to the BST test which will make the test sensitive to unequal variances in the CLEC and BST data, in the same way that the LCUG test is an adjustment to the pooled variance test.

4. Estimate of Variance in the Replicate Test – Because of concerns regarding the choice of variance estimator in the replicate estimate, we now use v_1 as the variance estimator, rather than the more conservative v_2. (Reference: Wolter, K. *Introduction to Variance Estimation,* 1985, Springer Verlag, New York.)

5. Jackknife Test – Because of concerns regarding the replicate technique, we have included an additional test which uses the jackknife approach. This, like the replicate variance estimate, uses the idea of subsample replication and a description can be found in Wolter's 1985 book.

6. When the Data are Uncorrelated – We have added a test of the hypothesis that the adjusted LCUG is suitable for a data set. If this null hypothesis is not rejected, then

the adjusted LCUG test procedure can be used. This is done using a two-tailed test of the null hypothesis H_0: "Modified LCUG test statistic" = "Adjusted Jackknife test." The data provided for the OSS Response Interval does not allow one to use the LCUG modified z test, nor the BST alternatives used on the Order Completion Interval and Maintenance Average Duration data sets. In the Interim Analysis Report we proposed using a modified t test that is based on time series analysis and generalized least squares estimation. This approach is still being used.

Based on the data that we have analyzed so far, Ernst & Young recommends that the Adjusted Jackknife Test described below be used on the aggregated data when the data are reported with enough detail. In cases where the data do not have sufficient detail, alternate approaches like that used for the OSS Response Interval should be used.

Detailed Descriptions

1. Trimming the Extreme Tails of the Distributions

   We have provided a more general trimming rule that trims the BST order completion interval data more severely than in the previous analysis. The completion interval distributions seen up to this point have been skewed, with an extreme tail in only one direction – namely large values. The revised trimming rule in this case is to trim the largest 10 CLEC cases. All BST observations greater than the remaining largest CLEC observation are then trimmed. For example, in the data for the August Order Completion Interval, the 11 largest CLEC observations have the values 24,26,26, 26, 26,27, 28,28,28,34, and 46 days. The 10 observations with values greater than 24 are trimmed from the CLEC data. All BST observations with values greater than 24 are removed from analysis; these trimmed values range in value from 25 to 189 days. This results in 0.22% of the BST data being trimmed and 0.06% of the CLEC data being trimmed.

   Only the Order Completion Interval has been trimmed in this way. The OSS Response Interval data are inappropriate for this type of trimming, and no trimming is needed for percent measurements. The Maintenance Average Duration data has been trimmed at 240 hours in the past, and we have continued to do this. We will investigate applying the new trimming approach on this data in the future.

2. Adjustment by Subclassification to Remove Bias

   Because the data are not the result of a designed experiment but come from an observational study, bias is a serious concern. The true means of the performance measure may differ across classes, defined by time, location, and type of service, and the distribution of the CLEC observations over these classes may differ from the distribution of the BST observations. In this case, under the null hypothesis of no favoritism, the simple difference of means is a biased estimate, and therefore the Type I error is not correct. Adjustment by subclassification is a frequently used devise for trying to reduce such bias. Weighted averages of the subclass means are compared, using the same weights for the BST cases and for the CLEC cases.

Under the null hypothesis of no favoritism, any definition of the weights, such that the weights add to one, results in an unbiased estimate of the difference. The choice of weights is then made to satisfy other properties of interest. Usually the criteria used for choosing the weights is to minimize the variance of the estimate. The original choice of weights, which adjust the BST observations to the distribution of the CLEC observations, was made because a) it was felt that the distribution of the CLEC's would be the distribution of interest, and b) because we believed that the variance of the estimate using these weights would generally be smaller than the variance of the estimate weighting the CLEC observations to the BST distribution.

Using the same notation as in the Interim Statistical Analysis Report, we have

n_{1j} = the number of BST cases in subclass j

n_1 = the total number of BST cases $= \sum_j n_{1j}$

$\bar{x}_{1j}$ = the mean of the BST cases in subclass j

$\bar{x}_1 = \frac{1}{n_1}\sum_j n_{1j}\bar{x}_{1j}$ = the overall mean of the BST cases

n_{2j} = the number of CLEC cases in subclass j

$\bar{x}_{2j}$ = the mean of the CLEC cases in subclass j

$\bar{x}_2 = \frac{1}{n_2}\sum_j n_{2j}\bar{x}_{2j}$ = the overall mean of the CLEC cases

The estimated difference in the means, adjusted to the CLEC distribution is calculated as

$$\frac{1}{n_2}\sum_j n_{2j}(\bar{x}_{1j} - \bar{x}_{2j})$$

The estimated difference in the means, adjusted to the BST distribution is calculated as

$$\frac{1}{n_1}\sum_j n_{1j}(\bar{x}_{1j} - \bar{x}_{2j})$$

To clarify some apparent misunderstandings, note that if in fact the distribution of the BST's is the same as the distribution of the CLEC's over these subclassifications, then either adjustment results in exactly the same calculation as the simple difference in the means. That is, you still get the correct estimate. In other words, the adjustment does not in any way "hurt" if in fact it is not needed; in this case, the calculation gives you the simple difference in means.

3. Making the BST test more sensitive to the possibility that the BST variance may be smaller than the CLEC variance.

   The original LCUG test modified the pooled variance test by replacing the pooled variance estimate with the estimate of the BST variance. In a similar manner, an adjustment has been made to the t-statistics calculated using the replicate method and the jackknife method which will increase the absolute size of the test statistic if the estimated BST variance is smaller than the estimated CLEC variance, assuming independence. As with the original LCUG test, the adjusted test statistic will be smaller (less significant) than the unadjusted test statistic when the estimated BST variance is larger than the estimated CLEC variance.

   In general terms, the original BST test statistic is multiplied by the ratio of the estimated standard error of the estimate of the difference (the numerator of the test statistic) under the assumption of independence, divided by the standard error estimate where the CLEC variance estimate is replaced by the BST variance estimate. The exact formula for this adjustment is given in the appendix.

   For the test using the replicate variance estimate, the original statistic for the test is still given on the Decision page and is labeled "REP". The test statistic with this adjustment for disparity in the variances is labeled "REP ADJ."

4. Estimate of Variance in the Replicate Test.

   In the notation of the Interim Statistical Analysis Report, the estimate of variance now used in the calculation of the Replicate t-test is

$$v_1 = \frac{1}{G}\frac{1}{(G-1)}\sum_g (\bar{d}_g - \bar{\bar{d}})^2 .$$

   Reference: Interim Statistical Analysis Report, p. B-8.

5. Jackknife Estimate and Test Statistic

   Another subsample replication technique, called the jackknife, has been included. The jackknife methodology is a broadly useful technique in cases such as this, where the form or the properties of the point estimate are not straightforward. This methodology is used, in general, for two purposes a) to reduce bias, and b) to estimate variance. (Reference: Wolter (1985), *Introduction to Variance Estimation*, Section 4.2.) Using a combination of the notation in Wolter and in the Interim Statistical Analysis Report, the following is a brief description of the jackknife method used here.

An estimator $\hat{D}$ is calculated from the full data set. In the case where the BST observations are adjusted to the CLEC, $\hat{D} = \frac{1}{n_2}\sum_j n_{2j}(\bar{x}_{1j} - \bar{x}_{2j})$. The observations are then partitioned into G groups. We use the replicates, as defined for the replicate estimate, as the groups for the jackknife test.

Let $\hat{D}_{(g)}$ denote the estimator of the same functional form as $\hat{D}$ but calculated from the observations ***removing*** the g[th] group. (This is in contrast to the replicate methodology where we calculated the replicate estimate by using only the observations in replicate g.) Then G pseudo-values are defined and used for calculating the mean and variance, where the gth pseudo-value is defined as

$$\hat{D}_g = G * \hat{D} - (G-1) * \hat{D}_{(g)}$$

The estimate of the mean is the mean of the pseudo-values, $\hat{\bar{D}} = \frac{1}{G}\sum_{g=1}^{G}\hat{D}_g$ and the estimate of the variance of $\hat{\bar{D}}$ is $v(\hat{\bar{D}})) = \frac{1}{G(G-1)}\sum_{g=1}^{G}(\hat{D}_g - \hat{\bar{D}})^2$.

The statistic t= $\frac{\hat{\bar{D}}}{\sqrt{v(\hat{\bar{D}})}}$ is distributed approximately as a Student's t with G-1 degrees of freedom. This is the test statistic recorded on the Decision page as the JACK test.

The adjusted jackknife, referred to on the Decision Page as JACK ADJ, is this t-statistic multiplied by the adjustment factor for unequal variances, as described in (3).

6. When the Observations Appear to be Uncorrelated

We found with the data for the performance measure Order Completion Interval that the observations are not independent, but rather there appears to be a clustering effect, or a correlation between observations in the same location. However it appears that the observations for the Maintenance Average Duration, while having different distribution with respect to location may not be correlated. If that is true, then the adjusted or modified LCUG test is appropriate. We have therefore added a test of the hypothesis that the adjusted LCUG test is suitable for the data. If this null hypothesis is not rejected, then the adjusted LCUG test procedure can be used. If the null hypothesis is rejected, then the LCUG test is not appropriate and the BST test should be used.

A two-tailed test of the null hypothesis H_0: "Modified LCUG test statistic" = "Adjusted Jackknife test" is used. (The hypothesis test is made using the estimates with the BST data adjusted to the CLEC distribution.) This test is performed using a jackknife test. The general jackknife procedure, as described in (5), is applied but now the parameter of interest is not the difference between the BST means and the CLEC means. The parameter of interest is the LCUG test statistic minus the adjusted jackknife test statistic.

Equations

This section provides the equations used for the calculations on the Descriptive Page and the Decision Page of the performance measure analysis reports. The statistical tests used are based on the difference between the mean of the BST and the mean of the CLEC cases. Proportions are means, so these equations also apply to tests based on the difference between proportions or rates.

Notation:

n_1 = the number of BST cases

n_{1j} = the number of BST cases in subclass j

x_{1i} = the value of the performance measure for the i^{th} BST observation

$\bar{x}_1$ = the mean of the BST observations

$\bar{x}_{1j}$ = the mean of the BST observations in subclass j

$$s_1^2 = \frac{\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2}{(n_1 - 1)}$$

Similar notation using the subscript 2 is used to denote the values for the CLEC cases, that is

n_2 = the number of CLEC cases, etc

Adjusted to CLEC

In this case the BST observations are adjusted to the CLEC distribution over the subclasses. The adjusted or weighted mean for the BST cases is

$$\bar{x}_{1w} = \frac{1}{n_2}\sum_j n_{2j}\bar{x}_{1j} = \frac{\sum_j w_{1j}\sum_{i=1}^{n_{1j}} x_{1i}}{\sum_j w_{1j}n_{1j}} \quad \text{where } w_{1j} = \frac{n_{2j}}{n_{1j}} \tag{E.1}$$

and the weighted estimate of the BST variance is

$$s_{1w}^2 = \frac{\sum_j w_{1j}\sum_{i=1}^{n_{1j}}(x_{1i} - \bar{x}_{1w})^2}{\sum_j w_{1j}n_{1j} - 1} \tag{E.2}$$

The estimate of the difference in means is

$$\bar{x}_{1w} - \bar{x}_2 \tag{E.3}$$

and the LCUG test, adjusted to the CLEC's, is

$$\frac{\bar{x}_{1w} - \bar{x}_2}{s_{1w}\sqrt{c_1 + \frac{1}{n_2}}} \quad \text{where } c_1 = \frac{\sum_j w_{1j}^2 n_{1j}}{(\sum_j w_{1j}n_{1j})^2} \tag{E.4}.$$

The replicate test has been described previously and the jackknife test was described in a previous section. The estimate being calculated in each is the difference in means as in (A.3).

To increase the sensitivity of the BST test to inequality of variances, the jackknife test, and the replicate test, are multiplied by an adjustment factor. Under the assumption that the BST observations are independent and identically distributed (IID) and the CLEC observations are IID, but allowing that the BST and the CLEC observations may have different variances, the expected value of the standard error used in the denominator of the jackknife and replicate tests is

$$\sqrt{c_1\sigma_1^2 + \frac{\sigma_2^2}{n_2}}$$

Therefore to make an adjustment similar to the LCUG adjustment to the pooled variance test, we multiply the jackknife (and replicate test) by

$$\frac{\sqrt{c_1 s_{1w}^2 + \frac{s_2^2}{n_2}}}{s_{1w}\sqrt{c_1 + \frac{1}{n_2}}} \tag{E.5}$$

where c_1 is defined in (A.4).

Adjusted to BST

In this case, the CLEC observations are weighted to the distribution of the BST cases. The LCUG test adjusted to the BST is calculated as

$$\frac{\bar{x}_1 - \bar{x}_{2w}}{s_1\sqrt{\frac{1}{n_1} + c_2}}$$

$$\text{where } c_2 = \frac{\sum_j w_{2j}^2 n_{2j}}{(\sum_j w_{2j} n_{2j})^2}$$

$$w_{2j} = {}^{n_{1j}}\!\big/_{n_{2j}} \text{ ,and}$$

$$\bar{x}_{2w} = \frac{1}{n_1}\sum_j n_{1j}\bar{x}_{2j} = \frac{\sum_j w_{2j}\sum_{i=1}^{n_{2j}} x_{2i}}{\sum_j w_{2j} n_{2j}}.$$

The adjustment factor to the jackknife and replicate test in this case is

$$\frac{\sqrt{\frac{s_1^2}{n_1} + c_2 s_{2w}^2}}{s_1\sqrt{\frac{1}{n_1} + c_2}}.$$

BELLSOUTH ACTION ITEMS ITEM 16-13
Balancing Type I and Type II Errors

This note is the first of a set of discussions concerning the types of error that are present in hypothesis testing. We first address the issue of balancing the risk of Type I and Type II errors. The important issue of comparing the probability of these errors occurring based on the LCUG modified z test and the alternative test proposed by BellSouth Telecommunications, Inc. (BST) will be addressed at a later date.

Type I error is the error that occurs when the null hypothesis that there is no favoritism on the part of BellSouth is true and we reject it. If we have correctly specified the null distribution, it is controlled directly by the specification of the critical value where we decide to either accept or reject the null hypothesis of no favoritism. Type II error is the error that occurs when the null hypothesis of no favoritism is false but we mistakenly accept anyway. Type II error is not controlled directly but decreases as the sample size increases.

In a controlled experimental study, where the sample sizes are relatively small, it is generally desirable to control the Type I error closely to avoid making a conclusion that there is a difference when, in fact, there is none. The probability of a Type II error is not directly controlled but is determined by the distance between the null hypothesis and the alternative and the sample size. Thus, there is some kind of balance between Type I and Type II errors with Type I error usually controlled more closely.

In Figure 1 below, the distribution assuming the null hypothesis is true is labeled H_0 and the distribution assuming a particular alternative difference between BellSouth and CLEC means is true is labeled H_a. The probability of a Type I error is the area under the null distribution to the left of the test critical value *c*. This region is labeled α. The critical value *c* determines the point beyond which an observed z-value is extreme enough to conclude that BellSouth is favoring itself. This is the decision rule that guides our determination of statistical significance. If, in fact and unknown to us, the alternative distribution is actually the true distribution, we still declare any test statistic that falls to the right of c to be significant. If it falls to the left of *c*, it is not significant. With respect to the alternative distribution, we can see that the area to the left of *c* will lead to an acceptance of the null hypothesis, even though, in this case, it is not true. The probability of a Type II error, incorrectly accepting the null hypothesis for a given correct alternative value, is labeled β on the graphic. Both α and β can be determined for specified null and alternative distributions.

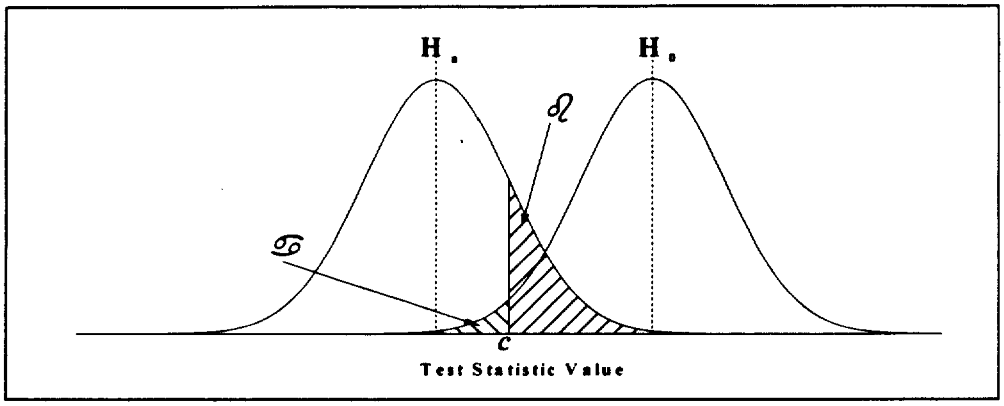


Figure 1

For the purpose of discussion, we consider the modified z test statistic proposed by the LCUG for testing the hypothesis of no favoritism. Let

n_1 = the number of BST observations,

n_2 = the number of CLEC observations,

$\overline{X}_1$ = the average performance measure value of the BST observations,

$\overline{X}_2$ = the average performance measure value of the BST observations, and

s_1 = the sample standard deviation of the BST observations.

The modified z statistics is

$$z = \frac{\overline{X}_1 - \overline{X}_2}{s_1 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

One interpretation of the null hypothesis that there is no favoritism on the part of BellSouth is that the true means of the BST and CLEC performance measures are equal, as well as the true standard deviations. Suppose that all the observations are independent, and the null hypothesis of no favoritism is true. If the number of BST and CLEC observations are sufficiently large then z has a standard normal distribution. A critical value for the test, given a value for the Type I error α, can be found from a table of the standard normal distribution, or through the use of statistical computer software.

To determine β, we must specifically state the alternative hypothesis. One way to do this is to assume that the true CLEC mean, μ_2, is actually larger than the true BST mean, μ_1, by some fraction of the true BST standard deviation, σ. That is,

$$H_a: \mu_2 - \mu_1 = f\sigma, \qquad f > 0.$$

It can be shown that the probability of a Type II error is given by the area under the standard normal density curve to the right of the value

$$c+\frac{f\sigma}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=c+\frac{f\sigma}{SE_\sigma(n_1,n_2)}. \qquad (1)$$

$SE_\sigma(n_1,n_2)$ denotes the standard error of the mean difference estimator $\overline{X}_1-\overline{X}_2$. The functional notation is used to emphasis the fact that for a fixed value of σ, the standard error varies as the number of observations for BST and CLEC varies.

Figure 2 shows graphs of the probability of a Type II error, β, versus the standard error of the mean difference estimator for $\alpha = 0.05$ ($c = -1.645$) and $f = 0.05$, 0.1, and 0.2. Notice that as the BST sample size, the CLEC sample size, or both sample sizes increase, the probability of a Type II error decreases.

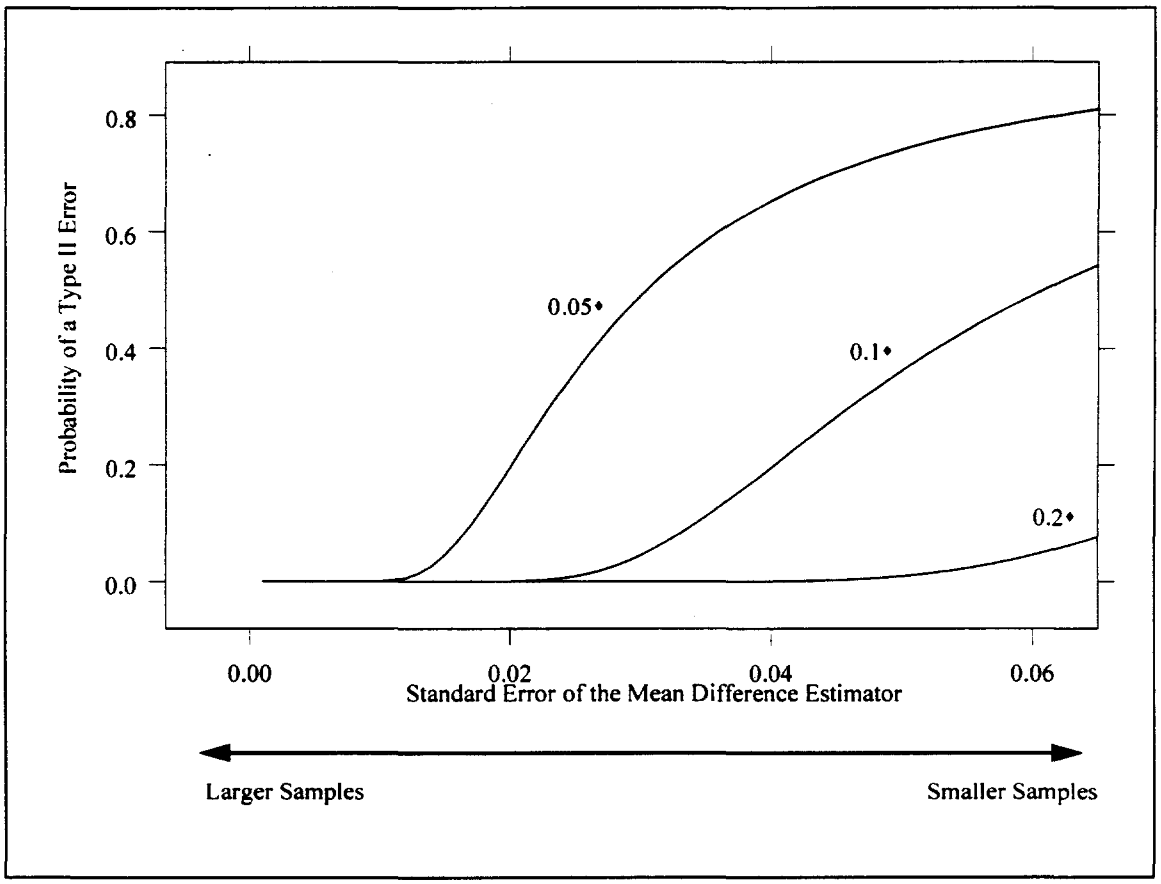


Figure 2: Probability of a Type II Error vs. Standard Error of the Mean Difference Estimator when $\alpha = 0.05$ and the mean difference of the alternative hypothesis is 0.05σ, 0.1σ, or 0.2σ.

In an observational study, where sample sizes are free to vary and may become very large, the balance between Type I and Type II errors can be reversed, with the Type I error risk remaining at a specified level (usually .05 or .01) and the risk of Type II error becoming very tiny. When that happens, we are much more likely to falsely reject a true null hypothesis of parity than we are to falsely accept an incorrect null hypothesis of parity.

To explore this further, suppose that the number of CLEC observations is some fixed proportion of the number of BST observations, that is, $n_2 = p \cdot n_1$ where $p > 0$. Then (1) can be rewritten as

$$c + \frac{f}{\sqrt{\frac{1}{n_1}\left(1 + \frac{1}{p}\right)}}.$$

Figure 3 shows graphs of the probability of a Type II error, β, versus the Number of BST Observations for $\alpha = 0.05$ ($c = -1.645$), $f = 0.1$, and $p = 0.05$, 0.04, and 0.03. Notice that β drops below 0.05, the value of α, when n_1 is approximately 23,000, 28,000 and 37,000 observations for the respective proportions p.

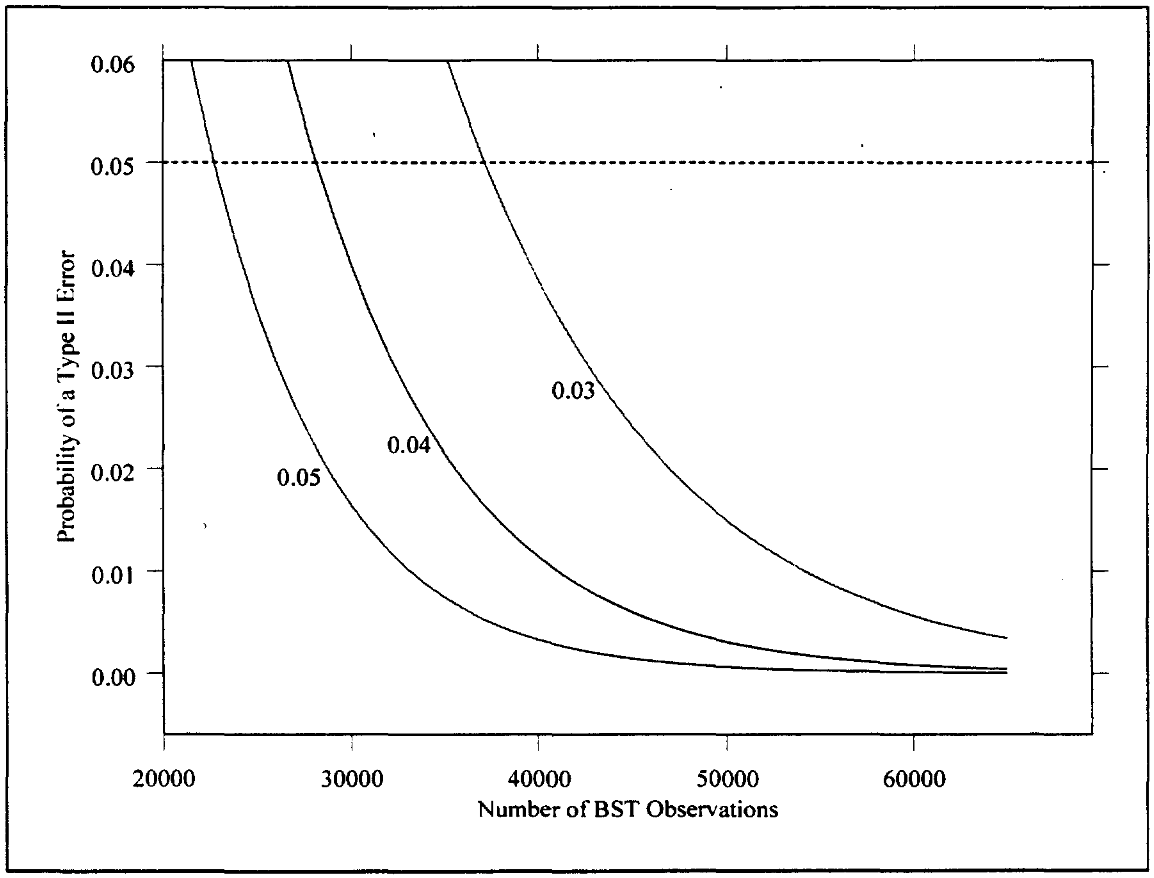


Figure 3: Probability of a Type II Error vs. Number of BST Observations when $\alpha = 0.05$, the mean difference of the alternative hypothesis is 0.1σ and the number of CLEC observations is 0.03, 0.04, or 0.05 times the number of BST observations.

Figure 3 is representative of situations that are possible for the BST/CLEC performance measure data that has been studied. There are many examples where BST has a very large number of observations with the proportion of CLEC observations in the range from 0.04 to 0.05. In these cases, the probability of a Type II error is much smaller than 0.05, the preset probability of a Type I error. To keep a balance between the two types of error, α should be lowered.

There are others issues as well that need to be considered. In an experimental design, the issue of materiality is addressed up front at the design stage in choosing the sample size needed to detect a given difference. This addressing of materiality or business impact often does not occur in the planning stages of an observational study like the BellSouth to CLEC performance comparison. However, it should be addressed in developing the rules that guide a decision of no favoritism or favoritism.

The issue here is not only one of keeping the **risk** of Type I and Type II error in balance; it is, more importantly, an issue of keeping the **costs** of Type I and Type II errors in balance. The cost to BellSouth of spending time and money to pursue the causes of false positives must be balanced against the cost to the CLECs of potential customer loss. Both costs should be explicitly considered. Simulation studies can be done to determine the sample size needed to keep the costs and risks of Type I and Type II errors in balance.

If the result of a statistical test is significant, it should then be compared to a materiality standard to determine whether favoritism exists. If a difference is not statistically significant, even if large enough to exceed the materiality standard, no favoritism exists. In other words, the measured difference must be both accurate enough to trust and large enough to have a business impact.